CSE574 Introduction to Machine Learning

Machine Learning: Notation and Definitions

Jue Guo

University at Buffalo

January 24, 2024

Outline

Notation

Let’s breifly revisit the mathematical notation we all learned at school.

Data Structure

A **scalar** is a simple numerical value, like 15 or -3.25 . Variables or constants that take scalar values are denoted by an italic letter, like x or a.
A **vector** is an ordered list of scalar values, called attributes. We denote a vector as a bold character, for example, **x** or **w**.

- Vectors can be visualized as arrows that point to some directions as well as points in a multi-dimensional space.

Illustrations of three two-dimensional vectors, $\mathbf{a} = [2, 3]$, $\mathbf{b} = [-2, 5]$, and $\mathbf{c} = [1, 0]$ are given in the figure.

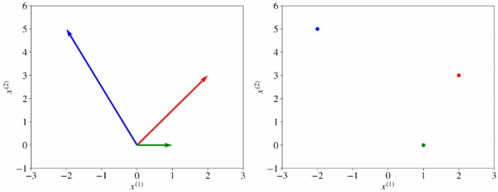

Figure: Three vectors visualized as directions and as points.

We denote an attribute of a vector as an italic value with an index, like this: $w^{(j)}$ or $x^{(j)}$. The index j denotes a specific **dimension** of the vector, the position of an attribute in the list. For instance, in the vector a shown in red in the figure, $a^{(1)} = 2$ and $a^{(2)} = 3$.

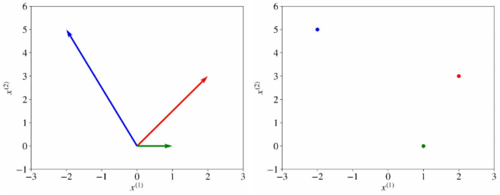

Figure: Three vectors visualized as directions and as points.

The notation $x^{(j)}$ should not be confused with the power operator, such as the 2 in x^2 (squared) or 3 in x^3 (cubed). If we want to apply a power operator, say square, to an indexed attribute of a vector, we write like this: $\left(x^{(j)}\right)^2$.

A variable can have two or more indices, like this: $x_i^{(j)}$ or like this $x_{i,j}^{(k)}$. For example, in neural networks, we denote as $x_{l,u}^{(j)}$ the input feature j of unit u in layer l.

A **matrix** is a rectangular array of numbers arranged in rows and columns. Below is an example of a matrix with two rows and three columns,

$$\begin{bmatrix} 2 & 4 & -3 \\ 21 & -6 & -1 \end{bmatrix}$$

Matrices are denoted with bold capital letters, such as $\mathbf{A}$ or $\mathbf{W}$.
A **set** is an unordered collection of unique elements.

- We denote a set as a calligraphic capital character, for example, $\mathcal{S}$.

A set of numbers can be finite (include a fixed amount of values).

- In this case, it is denoted using accolades, for example, $\{1, 3, 18, 23, 235\}$ or $\{x_1, x_2, x_3, x_4, \ldots, x_n\}$.

A set can be infinite and include all values in some interval.

- If a set includes all values between a and b, including a and b, it is denoted using brackets as $[a, b]$. If the set doesn't include the values a and b, such a set is denoted using parentheses like this: (a, b).
- For example, the set $[0, 1]$ includes such values as $0, 0.0001, 0.25, 0.784, 0.9995$, and 1.0. A special set denoted $\mathbb{R}$ includes all numbers from minus infinity to plus infinity.

When an element x belongs to a set $\mathcal{S}$, we write $x \in \mathcal{S}$.

- We can obtain a new set $\mathcal{S}_3$ as an **intersection** of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$. In this case, we write $\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cap \mathcal{S}_2$. For example $\{1, 3, 5, 8\} \cap \{1, 8, 4\}$ gives the new set $\{1, 8\}$.
- We can obtain a new set $\mathcal{S}_3$ as a **union** of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$. In this case, we write $\mathcal{S}_3 \leftarrow \mathcal{S}_1 \cup \mathcal{S}_2$. For example $\{1, 3, 5, 8\} \cup \{1, 8, 4\}$ gives the new set $\{1, 3, 4, 5, 8\}$.

Capital Sigma Notation

The summation over a collection $\mathcal{X} = \{x_1, x_2, \ldots, x_{n-1}, x_n\}$ or over the attributes of a vector $\mathbf{x} = \left[x^{(1)}, x^{(2)}, \ldots, x^{(m-1)}, x^{(m)}\right]$ is denoted like this:

$$\sum_{i=1}^{n} x_i \stackrel{\text{def}}{=} x_1 + x_2 + \ldots + x_{n-1} + x_n,$$

$$\text{or else: } \sum_{j=1}^{m} x^{(j)} \stackrel{\text{def}}{=} x^{(1)} + x^{(2)} + \ldots + x^{(m-1)} + x^{(m)}$$

The notation $\stackrel{\text{def}}{=}$ means "is defined as".

Capital Pi Notation

A notation analogous to capital sigma is the capital pi notation. It denotes a product of elements in a collection or attributes of a vector:

$$\prod_{i=1}^{n} x_i \stackrel{\text{def}}{=} x_1 \cdot x_2 \cdot \ldots \cdot x_{n-1} \cdot x_n$$

where $a \cdot b$ means a multiplied by b. Where possible, we omit $\cdot$ to simplify the notation, so ab also means a multiplied by b.

Operations on Sets

A derived set creation operator looks like this: $\mathcal{S}' \leftarrow \{x^2 \mid x \in \mathcal{S}, x > 3\}$.

- This notation means that we create a new set $\mathcal{S}'$ by putting into it x squared such that x is in $\mathcal{S}$, and x is greater than 3.

The cardinality operator $|\mathcal{S}|$ returns the number of elements in set $\mathcal{S}$.

Operations on Vectors

- The sum of two vectors $\mathbf{x} + \mathbf{z}$ is defined as the vector $\left[x^{(1)} + z^{(1)}, x^{(2)} + z^{(2)}, \ldots, x^{(m)} + z^{(m)}\right]$.
- The difference of two vectors $\mathbf{x} - \mathbf{z}$ is defined as $\left[x^{(1)} - z^{(1)}, x^{(2)} - z^{(2)}, \ldots, x^{(m)} - z^{(m)}\right]$.
- A vector multiplied by a scalar is a vector. For example $\mathbf{x}c \stackrel{\text{def}}{=} \left[cx^{(1)}, cx^{(2)}, \ldots, cx^{(m)}\right]$.
- A dot-product of two vectors is a scalar.
  - For example, $\mathbf{wx} \stackrel{\text{def}}{=} \sum_{i=1}^{m} w^{(i)}x^{(i)}$. The dot-product is denoted as $\mathbf{w} \cdot \mathbf{x}$. The two vectors must be of the same dimensionality. Otherwise, the dot-product is undefined.
- The multiplication of a matrix $\mathbf{W}$ by a vector $\mathbf{x}$ results in another vector. Let our matrix be,

$$\mathbf{W} = \begin{bmatrix} w^{(1,1)} & w^{(1,2)} & w^{(1,3)} \\ w^{(2,1)} & w^{(2,2)} & w^{(2,3)} \end{bmatrix}$$

When vectors participate in operations on matrices, a vector is by default represented as a matrix with one column. When the vector is on the right of the matrix, it remains a column vector. We can only multiply a matrix by vector if the vector has the same number of rows as the number of columns in the matrix. Let our vector be $\mathbf{x} \stackrel{\text{def}}{=} \left[x^{(1)}, x^{(2)}, x^{(3)}\right]$. Then $\mathbf{W}\mathbf{x}$ is a two-dimensional vector defined as,

$$\begin{aligned}
\mathbf{W}\mathbf{x} &= \begin{bmatrix} w^{(1,1)} & w^{(1,2)} & w^{(1,3)} \\ w^{(2,1)} & w^{(2,2)} & w^{(2,3)} \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ x^{(3)} \end{bmatrix} \\
&\stackrel{\text{def}}{=} \begin{bmatrix} w^{(1,1)}x^{(1)} + w^{(1,2)}x^{(2)} + w^{(1,3)}x^{(3)} \\ w^{(2,1)}x^{(1)} + w^{(2,2)}x^{(2)} + w^{(2,3)}x^{(3)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{w}^{(1)}\mathbf{x} \\ \mathbf{w}^{(2)}\mathbf{x} \end{bmatrix}
\end{aligned}$$

If our matrix had, say, five rows, the result of the product would be a five-dimensional vector.

When the vector is on the left side of the matrix in the multiplication, then it has to be **transposed** before we multiply it by the matrix. The transpose of the vector $\mathbf{x}$ denoted as $\mathbf{x}^\top$ makes a row vector out of a column vector. Let's say,

$$\mathbf{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix}, \text{ then } \mathbf{x}^\top \stackrel{\text{def}}{=} \begin{bmatrix} x^{(1)} & x^{(2)} \end{bmatrix}$$

The multiplication of the vector $\mathbf{x}$ by the matrix $\mathbf{W}$ is given by $\mathbf{x}^\top\mathbf{W}$,

$$\begin{aligned} \mathbf{x}^\top\mathbf{W} &= \begin{bmatrix} x^{(1)} & x^{(2)} \end{bmatrix} \begin{bmatrix} w^{(1,1)} & w^{(1,2)} & w^{(1,3)} \\ w^{(2,1)} & w^{(2,2)} & w^{(2,3)} \end{bmatrix} \\ &\stackrel{\text{def}}{=} [w^{(1,1)}x^{(1)} + w^{(2,1)}x^{(2)}, w^{(1,2)}x^{(1)} + w^{(2,2)}x^{(2)}, w^{(1,3)}x^{(1)} \\ &+ w^{(2,3)}x^{(2)}] \end{aligned}$$

As you can see, we can only multiply a vector by a matrix if the vector has the same number of dimensions as the number of rows in the matrix.

Functions

A **function** is a relation that associates each element x of a set $\mathcal{X}$, the **domain** of the function, to a single element y of another set $\mathcal{Y}$, the **codomain** of the function. A function usually has a name.

- If the function is called f, this relation is denoted $y = f(x)$ (read f of x), the element x is the argument or input of the function, and y is the value of the function or the output. The symbol that is used for representing the input is the variable of the function (we often say that f is a function of the variable x).

We say that $f(x)$ has a **local minimum** at $x = c$ if $f(x) \geq f(c)$ for every x in some open interval around $x = c$.

- An **interval** is a set of real numbers with the property that any number that lies between two numbers in the set is also included in the set. An open interval does not include its endpoints and is denoted using parentheses. For example, $(0, 1)$ means "all numbers greater than 0 and less than 1 ".

The minimal value among all the local minima is called the **global minimum**.

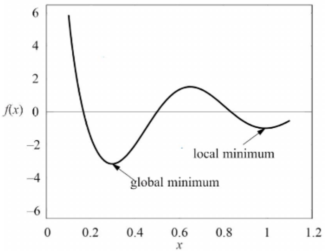

Figure: A local and a global minima of a function.

A vector function, denoted as $y = \boldsymbol{f}(x)$ is a function that returns a vector y. It can have a vector or a scalar argument.

Max and ArgMax

Given a set of values $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$,

- the operator $\max_{a \in \mathcal{A}} f(a)$ returns the highest value $f(a)$ for all elements in the set $\mathcal{A}$.
- the operator $\arg\max_{a \in \mathcal{A}} f(a)$ returns the element of the set $\mathcal{A}$ that maximizes $f(a)$.

Sometimes, when the set is implicit or infinite, we can write $\max_a f(a)$ or $\arg\max_a f(a)$. Operators min and arg min operate in a similar manner.

Example

1. **Set of Values**
   $\mathcal{A}$:$\{Alice : 12000, Bob : 15000, Charlie : 18000, Diana : 10000\}$
2. **Function** $f(a)$: This is the function mapping each salesperson to their sales figure. For instance, $f(Alice) = 12000$.
3. **Using Max**: $\max_{a \in \mathcal{A}} f(a)$: This will return the highest sales figure, which is \$18,000.
4. **Using Arg Max**: $\arg\max_{a \in \mathcal{A}} f(a)$: This will return the salesperson who achieved this highest sales figure, which is Charlie.